DEPARTMENT OF THE AIR FORCE

AIR UNIVERSITY

# AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GCS/ENG/88D-19

ENHANCED AUTONOMOUS FACE
RECOGNITION MACHINE

Volume I of II

THESIS

David D. Sander, B.S.
Captain, USAF

AFIT/GCS/ENG/88D-19



Approved for public release; Distribution unlimited

AFIT/GCS/ENG/88D-19

# ENHANCED AUTONOMOUS FACE RECOGNITION MACHINE

Volume I

Thesis

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Computer Engineering

David D. Sander, B.S.

Captain, USAF

December, 1988

Approved for public release; distribution unlimited

## Acknowledgements

First, to God I give thanks for the strength to persevere, and for the role you have played in my life. Without you, I never would have accomplished this task.

To my fiance [redacted] Thank you for your love and encouragement and for the patience you have shown towards me over the past year, when school work took priority. Now I am all yours.

To my advisor Dr Matthew Kabrisky. Many thanks for the knowledge you have passed on to me and for the advice and support you have given me on this thesis effort. Many thanks to my readers, Dr Frank Brown and Dr Steve Rogers, for the excellent comments and suggestions provided.

To all others, the staff of the Signal Processing Lab for their support, my subjects for taking time from a busy schedule and all others who lent support, I express sincere thanks.

Table of Contents

Volume I

Volume II

# List of Figures

# List of Tables

## Abstract

This thesis continues work on the Autonomous Face Recognition Machine developed at AFIT in 1985. There were two major changes made to the system. The set of features extracted from the face for use in the recognition process, was changed. A higher dimensioned vector taken from the two-dimensional Discrete Fourier Transform of the face, was used in hope of increasing the separation of templates stored in the data base. Further research is needed to determine whether this change is beneficial to the system. The second change was to the decision rule used in recognition. The decision making portion of the system was replaced by a back propagation neural network. While providing equivalent recognition capability, this change provides a constant recognition time independent of the number of subjects trained into the system.

# ENHANCED AUTONOMOUS FACE RECOGNITION MACHINE

## I. Introduction

### 1.1 Background

In 1985 Russel developed a face recognition system at AFIT (11). The system was based on the Cortical Thought Theory (CTT), which was presented by Richard Routh in a doctoral dissertation (9). The CTT proposes that the brain extracts information in the form of a two-dimensional vector or "gestalt". This gestalt is proposed as the only information passed to higher levels of the brain for processing (11:3-1,3-2).

The system was improved in 1986 by Smith (12). Smith added an algorithm to automatically locate a human face in a scene. This algorithm eliminated the need for human interaction in the face location process. However, the location algorithm was slow and recognition was somewhat degraded. The recognition capabilities of the system decreased because the face locater was only able to provide the internal features of the face for processing. Smith was able to show, however, that a computer could tell whether a face is present in a scene and then identify that face (12:6-1,6-2).

Recent improvements by Lambert have resulted in the current Autonomous Face Recognition Machine (AFRM) (5). Lambert developed elegant brightness and contrast normalization mechanisms and video preprocesses which greatly improved the windowing algorithms (which located faces). He also increased the speed of the system partly by rehosting it on a faster computer (5:3-1,3-2).

## 1.2 Problem Statement

The goal of my thesis effort is to improve the AFIT face recognition system to eliminate some of the problems of the current system. One of the improvements will be to find a new feature set. A higher dimensioned feature vector should improve the recognition capabilities of the system, by increasing the separation of the template vectors stored for each person. Note that this now violates the two dimensional rule of the CTT gestalt mechanism. With the addition of more faces to the data base, a new method (classifier) is also needed which will work faster than the current method in deciding whether the test vector matches one of the templates.

## 1.3 Assumptions

Assumptions from previous work which remain in effect are as follows:

1. The subject(s) are looking squarely at the camera (the head is not tilted or rotated).
2. The subject(s) are not wearing glasses.
3. The subject(s) have relaxed expressions (the face is not deliberately contorted)
4. Four pictures are sufficient to characterize a person in the database. (12:1-5)

## 1.4 Standards

Standards from previous work which are still in effect are as follows:

1. The AFRM should demonstrate "human like" classification of faces.(12:15)
2. Recognition performance of the AFRM must remain at least as good as that obtained by Russel. (12:1-5)

3. No operator interaction is allowed in the face location, windowing and recognition processes. (5:1-4)

4. The AFRM should be able to process scenes with a random, uncontrolled background. (5:1-3)

5. The AFRM must be able to process scenes with multiple faces in them. (5:14)

## 1.5 Scope

There are many areas of the AFRM which need improvement. Lambert suggested ways to improve the image processing, the face location and the recognition capabilities of the AFRM. This thesis effort will be limited to improvements in the recognition capabilities of the system. Emphasis will be on changing the feature set and implementing the decision rule with a neural net.

## 1.6 Approach/Methodology

The first step was to find a new feature set. The 2D Discrete Fourier Transform (2DDFT) was chosen as a means for generating the new feature set. The next step was to find a new classifier for making the recognition decision. Lambert recommended using a method with a constant recognition time no matter how many faces are stored in the system. A neural network was selected to replace the decision making portion of the AFRM since some neural network structures are constant-time processors in spite of their stored content size. These first two modifications were each tested separately. After testing the AFRM was modified to include both ideas. This new system was then compared to the original AFRM developed by Lambert.

## 1.7 Materials and Equipment

In order to perform the modifications of the system, access was needed to the source code for Lambert's AFRM. Access was needed to the video camera and the SMV2A micro VAX computer in the signal processing laboratory for running and testing the systems. Also additional disk storage space was needed for the additional sets of images required for testing the system.

## 1.8 Other Support

A great deal of help was needed from my fellow students. Volunteers were needed to be digitized for training and testing the system. Eight pictures of each person were taken hoping to get five faces that the system can locate. Four of these five are used in training the systems and one is used to test the systems.

## 1.9 Overview

The purpose of chapter two is to give the reader an overview of the previous face recognition work at AFIT. The primary focus is on the current AFRM developed by Lambert. Work done by Russel and Smith is also mentioned.

Chapter three discusses the use of neural networks and 2DDFTs in pattern recognition. It provides background of work done by Ruck using neural networks, and work done by O'Hair using 2DDFTs. This chapter gives some justification for the modifications AFRM.

Chapter four describes the modifications made to the AFRM and other work done. Chapter five provides results and chapter six provides recommendations and conclusions.

# II. Background of Previous Face Recognition Work

## 2.1 System Hardware Configuration

In 1987, Lambert ported the AFRM to a Microvax II (SMV2A). SMV2A has a 9MByte main memory, three 71MByte hard disk drives, and a TK50 tape drive. Installed in the SMV2A is an FG-100-Q Image Processing System. This system includes a video processing board with video memory, an RGB video monitor, and a software support library. The SMV2A provided the environment that Lambert felt he needed to make his enhancements to the AFRM (5:3-1).

## 2.2 Image Acquisition

The AFRM provides several methods for acquiring images for processing. These methods are provided as options from a menu on the system. New images can be acquired from the camera in one of two ways. The first method takes the image of a stationary subject. The second method uses a moving target indicator (MTI) algorithm to take the image of a moving subject. The MTI method reduces the amount of computational effort required to locate a face in the scene, by narrowing the search area for faces. A newly acquired scene may be stored to disk for future processing. Thus there is an option to load a scene from disk for processing. Lambert provided a useful program (Autotake.c) for use in acquiring test data. This program takes four images of a person and combines them into one scene which is stored to disk. This program provides quick acquisition of faces and helps conserve disk space.

## 2.3 Image Processing

Several types of image processing are used to help in the face recognition process. These include image sharpening, brightness normalization, scaling and contrast enhancement. The first step, however, is face location.

2.3.1 **Face Location.** In the original work by Russel, faces were located automatically and reasonably accurately provided that the subjects were placed in front of a plain white background. Russel's program also allowed the use of a keypad to shift the indices marking the features and edges of the face. This method provided great accuracy in face location, but was not completely realistic in an operational environment. It was decided that all human intervention should be eliminated to make the system autonomous. An algorithm was developed by Smith to locate the face automatically, but it was able only to locate the internal features of a face (5:2-9). Lambert made several improvements to the algorithm. A feature was added to use moving target detection to narrow the area searched for a face. Lambert also added the use of an ellipse drawn around the internal features of the face as an estimate of the edges of the face (5:3-3,3-11).

2.3.2 **Sharpening.** Before a scene is processed to locate faces; the operator may choose to sharpen the image. According to Lambert, sharpening of the image sometimes helps the face-location process. The image sharpening is performed by a call to one of the image processing subroutines provided in the library (5:C-6).

2.3.3 **Brightness Normalization.** Brightness normalization is performed on all faces before the face recognition process begins. This normalization process begins by calculating the average brightness of neighboring pixels. Each pixel is then given a new value according to the equation

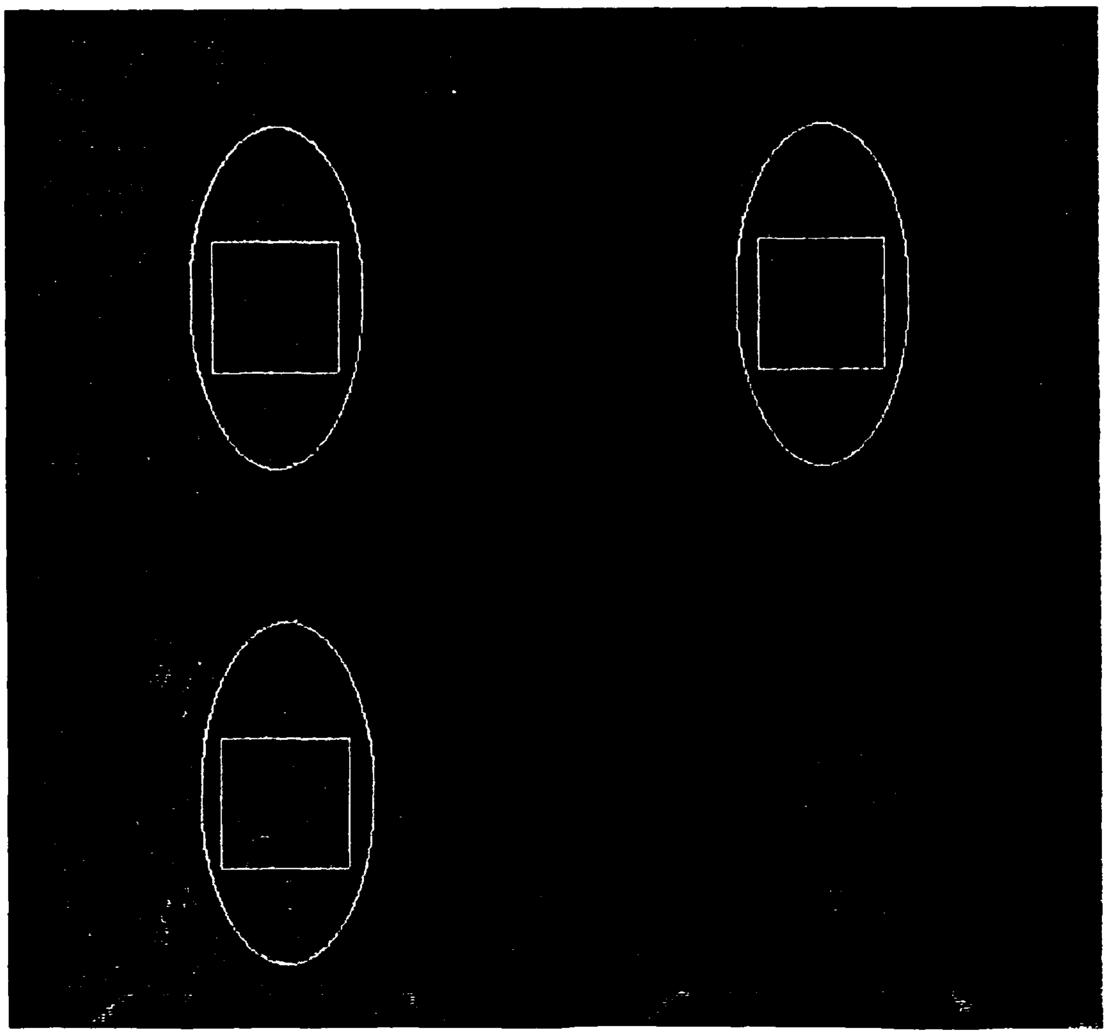

Figure 2-1. New Automatic Face Location Algorithm

$$\text{New pixel value} = 128 - \text{old pixel value} - \text{neighborhood average} \quad (1)$$

The average pixel value is calculated from neighboring pixels rather than from all pixels in the picture. This provides a local normalization which helps eliminate systematic changes in brightness. The main purpose for normalization is to help eliminate differences caused by lighting and camera settings (5:3-17).

2.3.4 **Scaling**. Faces which are smaller than 64 x 64 pixels are size- normalized by scaling them up to fill the 128 x 128 window. A graphics routine called zoom is used to double the size of these faces. Subsequent size normalization is also accomplished by scaling the results of the gestalt calculation.

2.3.5 **Contrast Enhancement.** Another process performed on the faces before calculating the gestalt is that of contrast enhancement. The contrast enhancement is performed by an ITEX library subroutine called HISTEQ. HISTEQ modifies the brightness values of a scene based on a histogram it generates from the pixel values. Each pixel in the image resulting from this process is then compared to a threshold. Any pixel with a value greater than 50 is set to 255. Thus the darkest pixels keep their value, but all others are set to the brightest value (5:3-22).

## 2.4 Windows

Russel first used the idea of creating windows which contain different portions of the face. This idea originated because the gestalt calculation can produce the same result on two similar faces which differ only in their width. This is due to the symmetry of faces and the nature of the calculation which is discussed later. By performing the calculation on

windows containing only portions of the face, the symmetry of the face is eliminated. Russel used six windows containing the following subsets of the face:

| Window | Features Included |
|---|---|
| 1 | Left side of head |
| 2 | Right side of head |
| 3 | Right side of head from top of eyes to chin |
| 4 | Right side of head from top of eyes to middle of mouth |
| 5 | Right side of head from tip of nose to chin |
| 6 | Right side of head from top of head to bottom of eyes |

Russel's other window choices are made based on experiments performed to see whether people can be recognized with portions of their faces blocked (11:4-29).

Smith used a different set of windows in his system. He was constrained by the fact that he only had the locations of the internal features of a face available. Since he used uncontrolled backgrounds, rather than uniform white backgrounds, he was unable to reliably determine facial boundaries such as the top of the head, the chin, and sides of the face. Because of this, his window set was completely different from Russel's.

Lambert, however, was able to use windows similar to those of Russel, because of the change he made in the face location algorithm. The ellipse which his algorithm puts around the internal features of the face provides a good estimate of the locations of the face edges. Lambert made a change in the windows to spread out the feature vectors in the decision space. Russel located the portion of the face in the window with respect to the upper left corner of the window. This created a clustering of the feature vectors. Lambert changed the location of the face portions. He moved them back to where they would normally be if the whole face were in the window. Lambert also changed which parts of the face are displayed in the windows. He made his changes based on the best performing windows of both Russel and Smith (5:3-33).


Figure 2-2. Russel's Window Set (11:B-32)

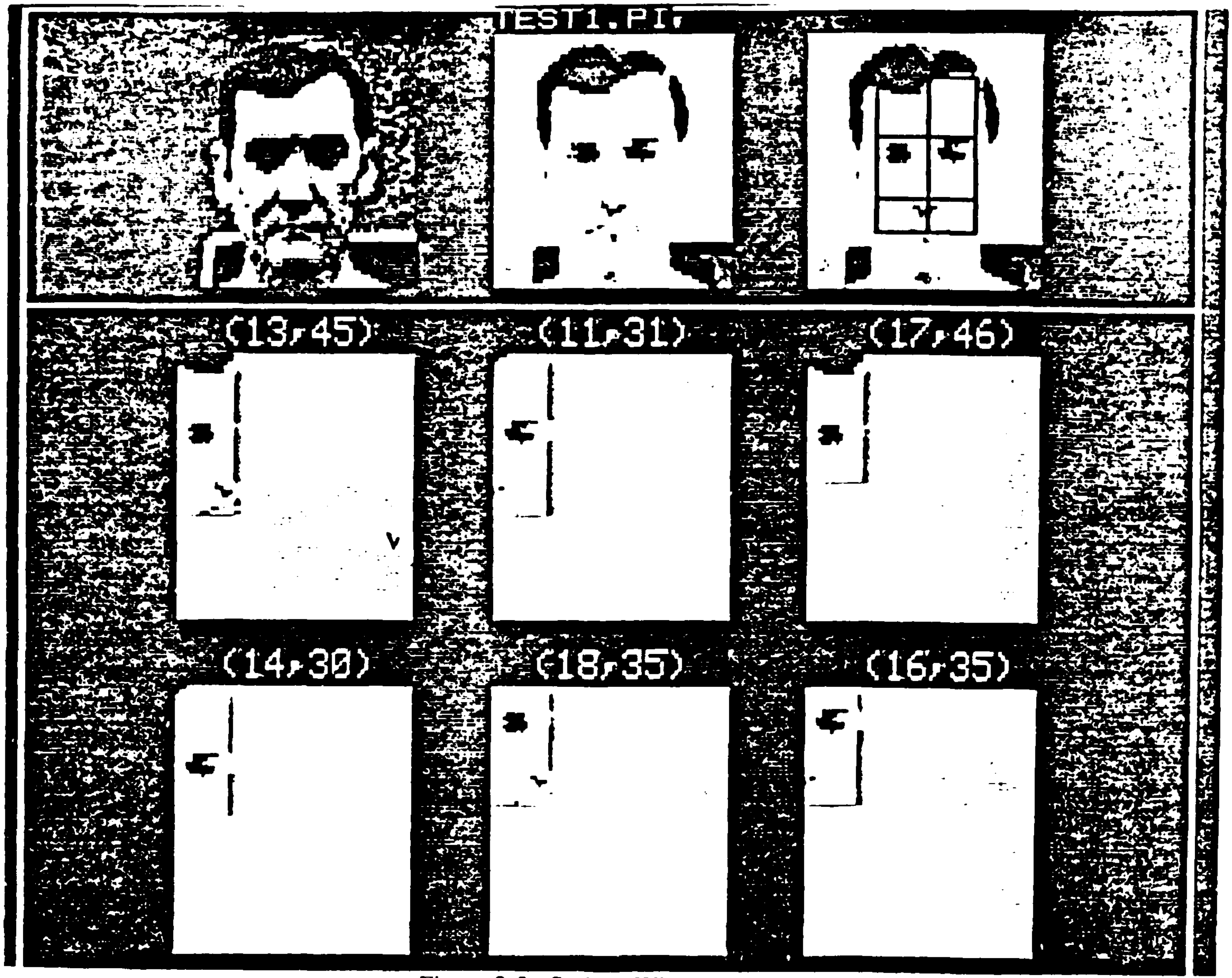


Figure 2-3. Smith's Window Set (12:3-32)

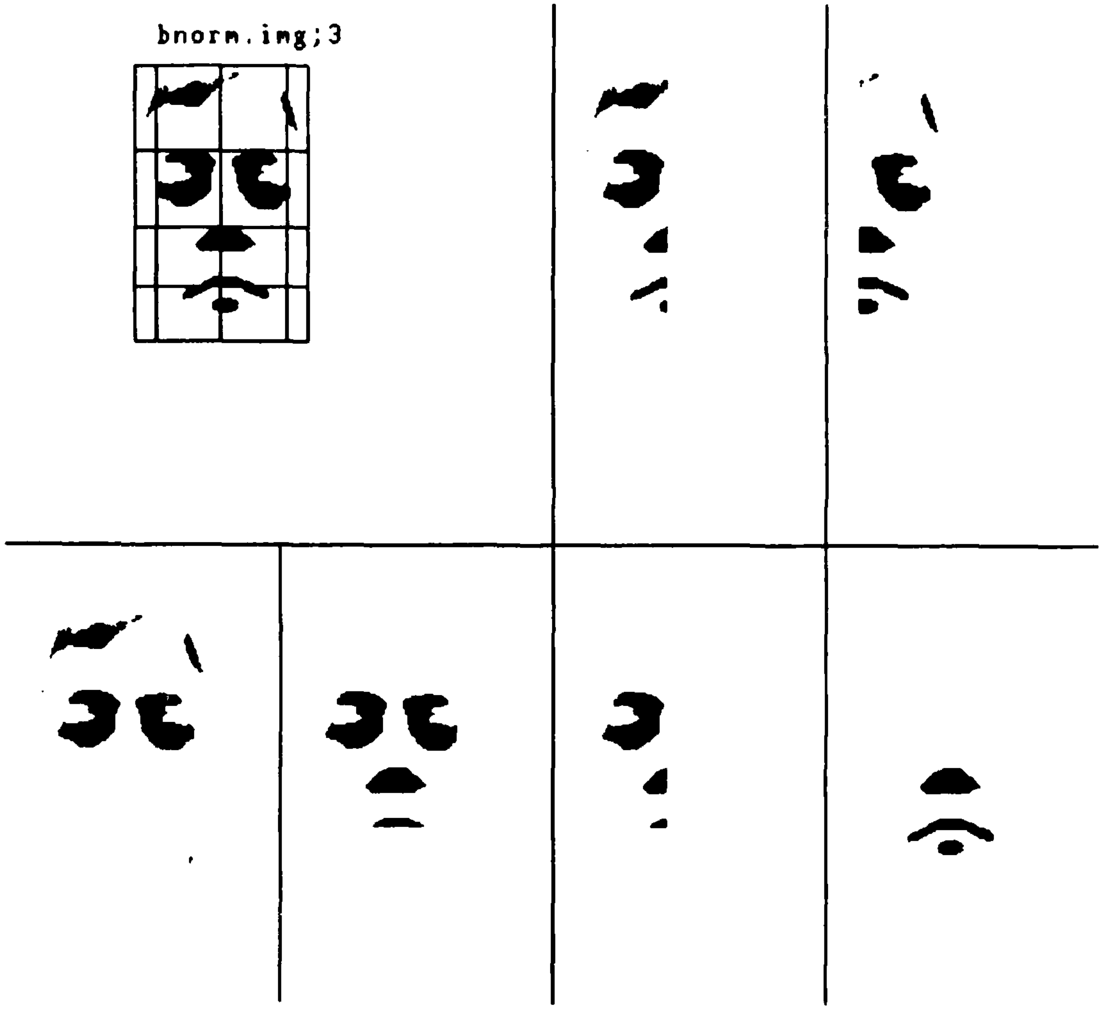


Figure 2-4. Lambert's Window Set

## 2.5 Gestalt Calculation

The gestalt calculation is used to produce a feature vector for each face. This feature vector is the basis for the recognition portion of the AFRM. The gestalt calculation is performed on each of the six windows. Two numbers are produced by each calculation. These numbers are combined to produce a feature vector for the face.

**2.5.1 1-D Gestalt Transform** There are several steps involved in the gestalt calculation (Figure 2-5). The first step in the gestalt transform on an array A of length L is to generate a gaussian distribution in an array G of size 2L-1. The result of the 1D gestalt transform is an array R of size L. Each element of R is created by.taking the dot-product of A and a portion of G. The portion of G used in the dot-product depends upon which element of R is being calculated. If element 1 is being calculated, then elements L through 2L-1 of G are used for the dot-product calculation. If element L is being calculated, then elements 1 through L of G are used (11:5-42,5-44).

**2.5.2 2-D Gestalt Transform** The 2D gestalt algorithm is based upon a similar algorithm for calculating a 2DDFT. The gestalt transform of each row is made, substituting the results back into the array. The gestalt transform is then made on each column of the array. The final result of the gestalt comes from the array coordinates of the largest value in the array as shown in Figure 2-6 (11:5-44,5-46).

## 2.6 Decision Mechanism

The heart of the AFRM is its decision mechanism. The decision mechanism used by the AFRM and by Russel is based on a standard pattern recognition technique. Using this technique, a template is generated for each face to be recognized. The distances between

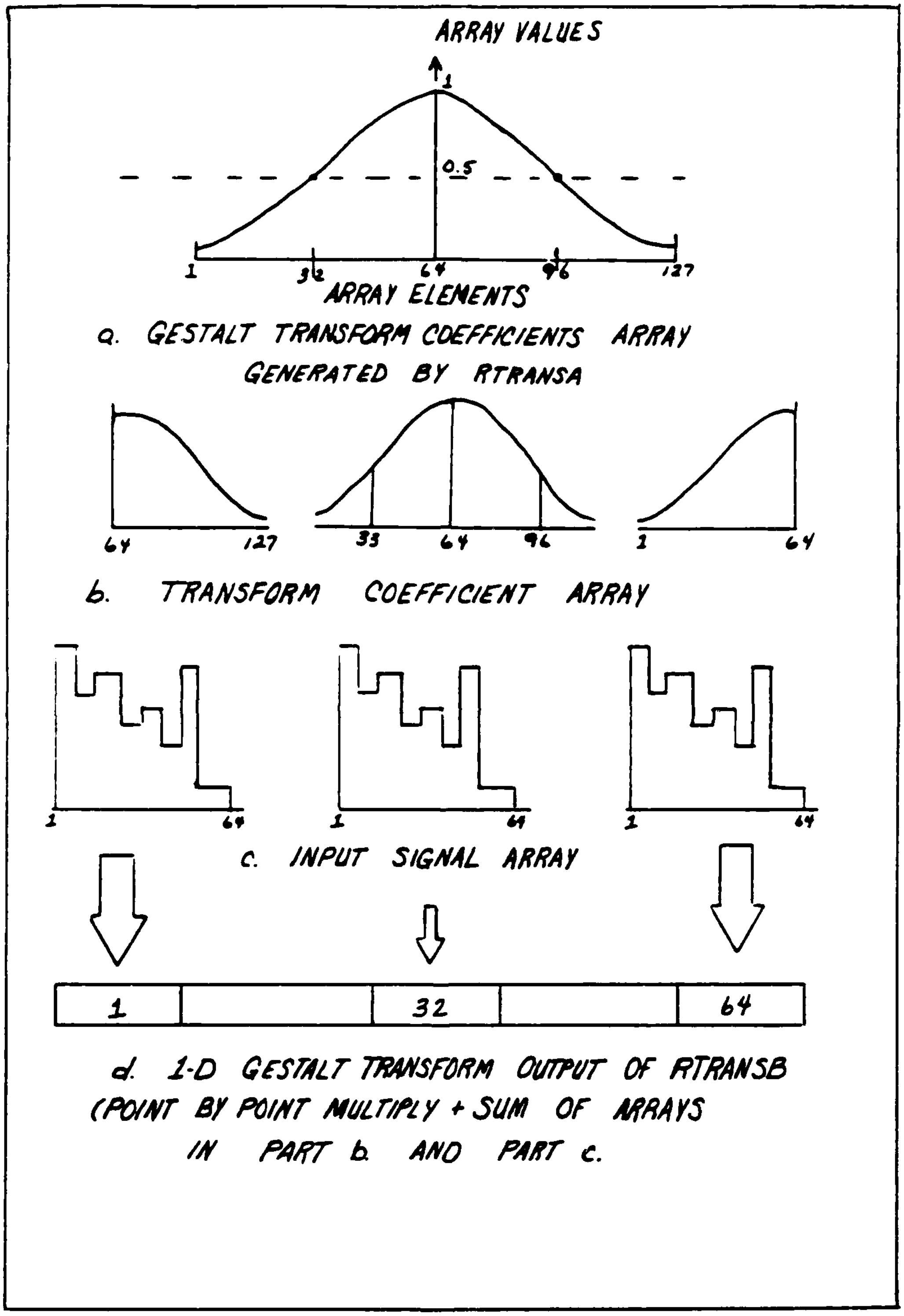


Figure 2-5 1-D Gestalt Transform(11:5-43)

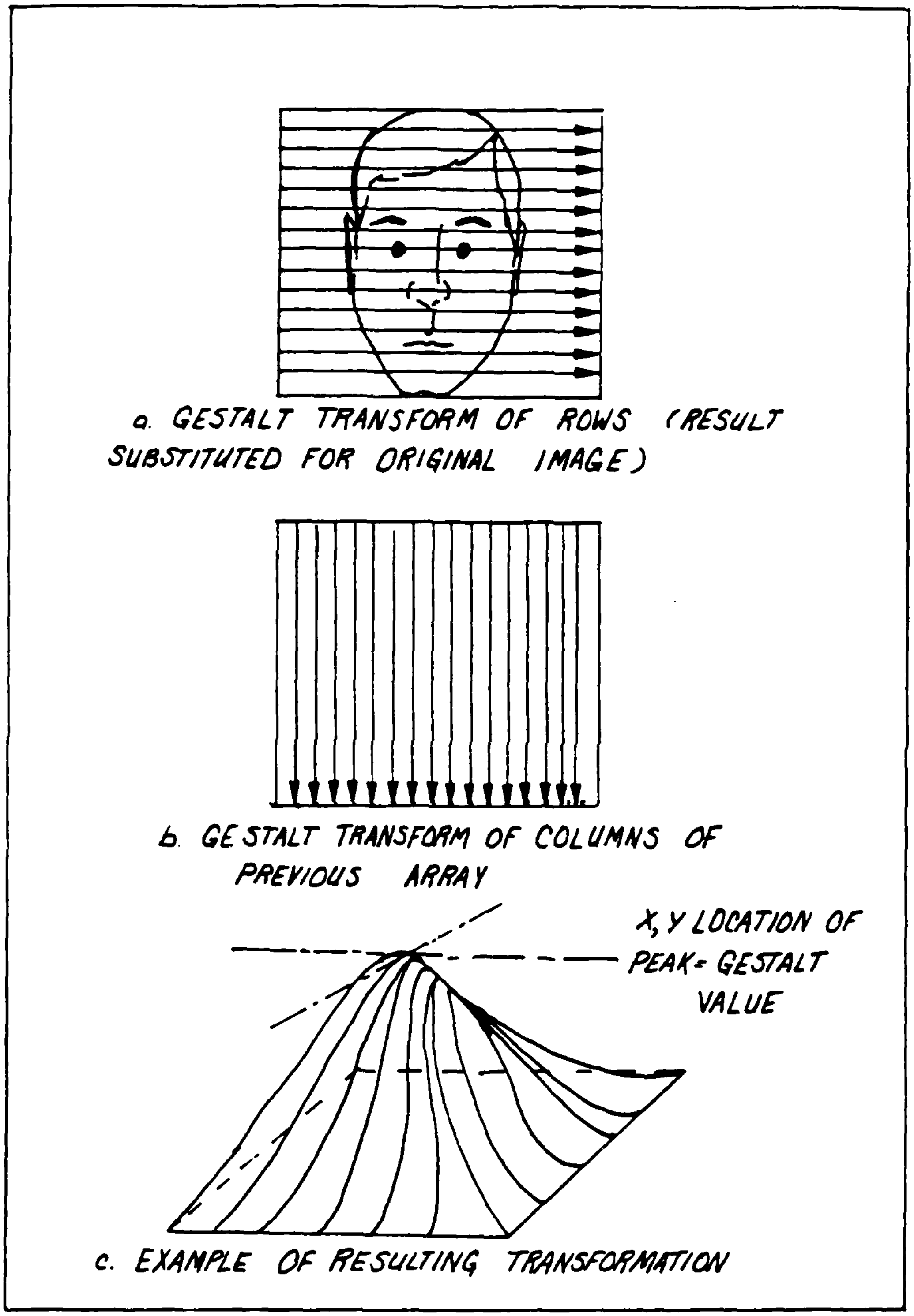


Figure 2-6. 2-D Gestalt Transform (11:5-45)

each template and an unknown test vector are calculated. The template which is closest to the unknown vector is chosen as the "recognized" template.

In the AFRM the template for a person is generated by averaging the vectors from the set of four faces that were used to train the system. The vector for an unknown face is then compared to each of these templates. The distances are then sorted. Pictures of the top three candidates are displayed on the monitor and all of the names are listed in rank order.

Lambert performed an experiment in which he used only one window in the recognition process. This was repeated for each of the six windows. He then compared the recognition capabilities of the window. Lambert found that the performance varied from window to window. This discovery gave him the idea that emphasis might be given to windows which perform better. Lambert made changes to the decision calculation resulting in the multiplication of the results of each window by window performance factors. Lambert was unable to show that this would improve the performance of the AFRM (5:3-40,5-10).

# III. Pattern Recognition Background

## 3.1 Neural Networks in Pattern Recognition

For years researchers have been trying to develop computer systems which are able to see and hear. This is a very complex problem and traditional attempts at solving this problem have been very computationally intensive. In traditional methods of pattern recognition a set of features is extracted from the input data. A subset of these feature vectors is saved to serve as a template. The feature vectors of inputs to be classified are compared to each of the template vectors by computing the distance between the vectors. The closest template is chosen as the match for the input. The more templates stored in the system the longer it takes to find a match (4:70).

Obviously the human brain does not work in this way. If it did, as people grew older and knew more faces it would take longer for them to recognize someone (8:52). The human brain has billions of neurons configured to perform such tasks in parallel. Recently more and more researchers have been attempting to model this capability using neural networks.

The concept of modeling the human brain is not new; some early ideas date back to the dawn of technology. One very interesting period began in the late 1950s with the Perceptron model of Rosenblatt. Although his concept was abandoned in the mid 1960s because of perceived weaknesses in the Perceptron model, recent discoveries have reopened this area of research. Nearly 2000 researchers attended the first international conference on neural networks in June 1987 where many models were discussed, including several sessions on Perceptrons (8:52).

A variety of neural network models have been developed. All use the concept of a neuron or node which has multiple inputs and a single output. A weight factor is associated with each input (See Figure 1). Inputs having positive weights are excitatory

inputs and those with negative weights are inhibitory inputs. Each input is multiplied by its weight. These products are added together to produce the total input for the node. The output of the node is based on some function of this input minus a threshold (6:5).

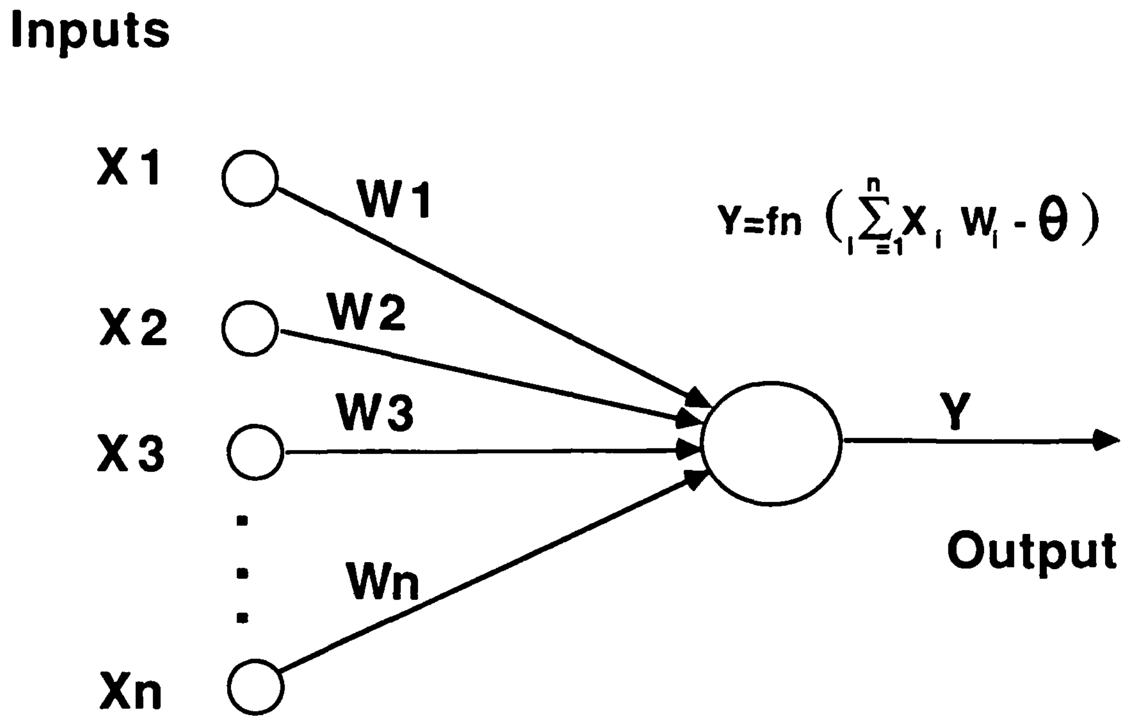


Figure 3-1. A Neural Network Node.

Most neural network models contain nodes like these. The networks differ primarily in the interconnection graphs of the nodes, the number and size of the layers, and the training methods used to change the input weights of the nodes.

In his article, Lippmann gives a good taxonomy of the networks which have been developed. His taxonomy is shown in Figure 3-2. He first divides the networks into those with binary inputs and those with continuous-valued inputs. Next the nets are divided into those with supervised training and those with unsupervised training. Those nets which use supervised training, such as Perceptrons, Hopfield nets and Hamming nets are generally

used as associative memories or classifiers. They are given the additional information of labels which specify the correct class for new input patterns during training. Those nets with unsupervised training, such as the Carpenter/Grossberg classifier and Kohonen's self-organizing feature maps are used to form clusters. The input to the network can then be classified according to the cluster in which the output falls (6:7). The abilities of all of these networks to classify inputs is what lends them to use in the area of pattern recognition. There are many examples of nets which have successfully solved problems in pattern recognition.

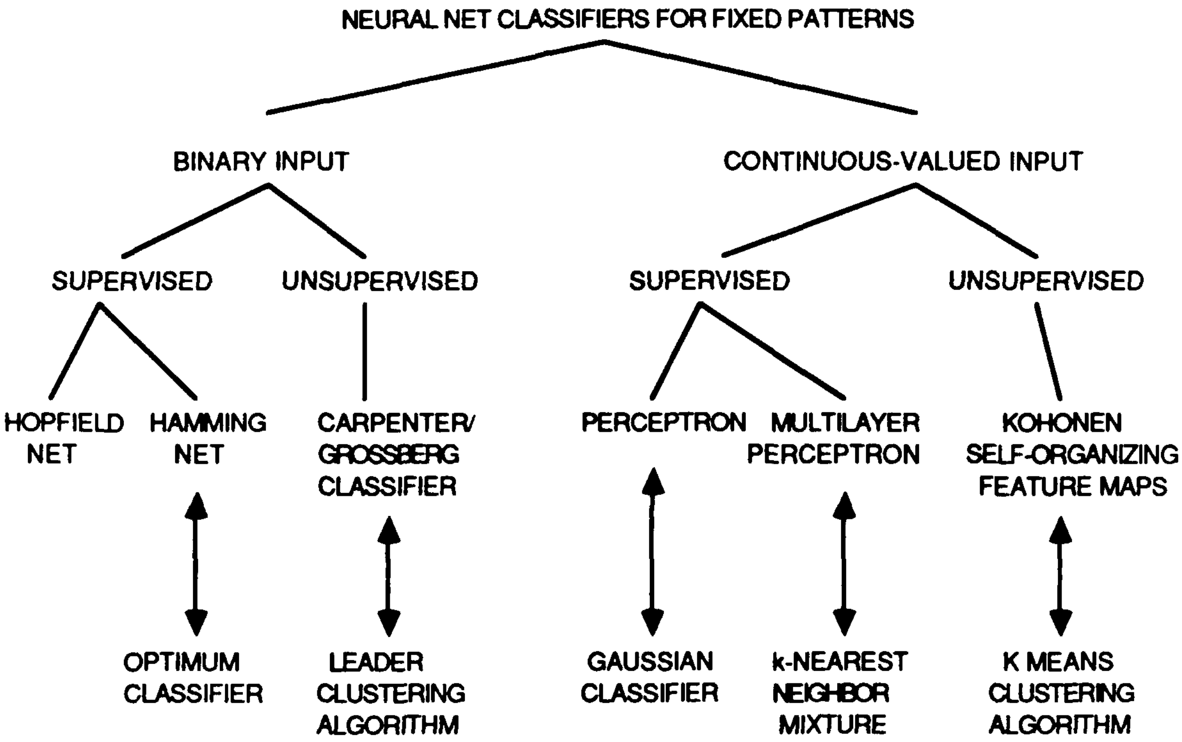


Figure 3-2. A taxonomy of six neural nets that can be used as classifiers. Classical algorithms which are most similar to the neural net models are listed along the bottom. (6:6)

One of the most popular networks used in recent research is the multilayer Perceptron. This network, based on early (1957) work which created the Perceptron, uses a "backward propagation" training algorithm. This algorithm has helped cause the reawakening of widespread interest in this field of research.

Work using these networks was recently completed by Dennis Ruck at AFIT. In his 1987 thesis Ruck describes how a multilayer Perceptron is used to classify feature vectors generated from sensor images of tanks, jeeps, POLs and trucks. He compares his network results with those generated by traditional pattern recognition techniques. Ruck states: "the multilayer Perceptron outperformed the statistical nearest-neighbor classifier in every test (10:4-30)". This would indicate that the multilayer Perceptron can be effectively used for automatic target recognition.

Another military application has been demonstrated in the recent work of Paul Gorman of the Allied-Signal Aerospace Technology Center and Terrence Sejnowski of Johns Hopkins University. The network they developed contained sixty inputs which were connected to a preprocessed sonar signal. The second layer, usually referred to as the hidden layer, consisted of twelve hidden nodes. The number of hidden nodes was determined by experimenting with hidden layers varying in size from 0-24 nodes. The output layer consisted of two nodes (3:76).

This network detected the difference in sonar signals produced by a rock and a metal cylinder. It achieved a classification accuracy as high as 100% when classifying inputs which were part of the training set and it correctly classified 90.4% of test samples not contained in the training set. It outperformed the nearest neighbor classifier which had an accuracy of 82.7%. The network performance was as good as that of the best trained human listeners (3:75).

T. A. Heppenheimer recently described some other work performed by Sejnowski. Along with Charles Rosenberg of Princeton University, Sejnowski developed a neural net

to produce speech directly from printed text. This network has seven letter inputs and fifty-five outputs as well as a hidden layer of nodes. The network scans text and outputs a phoneme to correspond to the middle letter of the seven letters presented to its input nodes. After each word it stops to compare its pronunciation of the word with the correct pronunciation given by the teacher. If the network is in error, it adjusts its weights (4:76,78).

When recounting one of his first successful overnight runs Sejnowski said:

> At first it gave a continuous stream of babble. It was just guessing; it had not learned to associate phonemes with the letters. As the run continued, it began to recognize constants and vowels. Then it discovered there were spaces between words. Now its stream of sound broke up into short bursts, separated by those spaces. At the end of the night it was reading quite understandably, correctly pronouncing some ninety-two percent of the letters (4:78)

The performance was good, but not as good as the best text-to-speech system, DECtalk, developed by Dennis Klatt and marketed by Digital Equipment Corporation, which is able to produce speech which is almost perfectly intelligible. However, it took the developers of DECtalk years to reach levels of performance that NETtalk achieved in just one day (4:78). The achievement of this network demonstrates some of the powerful "learning" capabilities of neural networks.

Some researchers in the field have developed their own network designs. One of these, which was not mentioned by Lippmann in his taxonomy, is the Neocognitron. The Neocognitron was developed in Japan by Sei Miyake and Kunihiko Fukushima. Their Neocognitron is a nine layer network and uses a special training method. The network has a two-dimensional array of inputs and a layer of ten outputs. Handwritten numerals are presented to the array of input nodes and the output indicates which of the ten digits was presented. This system has demonstrated the ability to correctly recognize handwritten numerals of various penmanship styles. This system can work even if the input is distorted

or cluttered with noise (2:832-833). This, however, is probably an example of a "toy" system in that the "problem" which it solves does not derive from real world data and is not extendible to the real world problem which presumably inspired it, namely the ability to read strings of printed text. It is well known that the ability to read single isolated letters cannot, in general, be extended to functioning reading systems, since there is *no reliable* way to separate single letters from actual text.

Most research in neural networks has been limited to software models running on single processor machines. The training of these models can take a significant length of time. However, these networks are now being developed on silicon chips in order to increase their speed. Bell Laboratories has developed some chips which can accept up to 256 bits of input and which can stabilize to a pattern within 500 nanoseconds as opposed to several seconds. This speed improvement should allow the training and testing of networks in a significantly shorter period of time (1:12).

As the literature suggests, neural networks show great promise in their application to the problem of pattern recognition. Although still in the research phase, neural networks implemented in hardware may soon appear as production systems capable of solving many difficult problems.

## 3.2 Discrete Fourier Transforms in Pattern Recognition

Finding an appropriate feature set is one of the most difficult tasks in the pattern recognition process. In most cases several feature sets are tried before one is found which works well. Any measurement can be tried as a feature but there is no reason to assume arbitrary features will be useful until they are tested. However, sometimes insight can be gained from looking at what feature set works well with similar data. One feature set has been found which works well with video data. In his 1984 thesis, Mark O'Hair filtered out

the lower three harmonics of the Two-Dimensional Discrete Fourier Transform (2DDFT) of an image for a feature set to recognize complete printed words (7:15).

Both real and imaginary components are produce by the 2DDFT of an image. The real and imaginary components are in separate arrays. The filtering of the lower three harmonics reduces these arrays to 7 x 7 as seen in Figure 3-3 and Figure 3-4.

| | | | | | | |
|---|---|---|---|---|---|---|
| -42.5 | 31.6 | 110.2 | 221.7 | 173.1 | 19.6 | -42.4 |
| -13.6 | -117.2 | -188.6 | 448.4 | 380.5 | -32.2 | -76.2 |
| -193.2 | -116.6 | -28.7 | -607.0 | -298.5 | -65.6 | -113.5 |
| -142.1 | -97.2 | 149.6 | 666.4 | 149.6 | -97.2 | -142.1 |
| -113.5 | -65.6 | 298.5 | 607.0 | -28.7 | -116.6 | -193.2 |
| -76.2 | -32.2 | 380.5 | 448.4 | -188.6 | -117.2 | -13.6 |
| -42.4 | 19.6 | 173.1 | 221.7 | 110.2 | 31.6 | -42.5 |

Figure 3-3. Real Components

| | | | | | | |
|---|---|---|---|---|---|---|
| -142.7 | -124.8 | -109.4 | -381.8 | -317.2 | 12.9 | 283.3 |
| -89.3 | -94.9 | -224.5 | -409.7 | -413.8 | 6.4 | 197.6 |
| -62.9 | -91.3 | -322.7 | -236.1 | -483.6 | -34.9 | 128.4 |
| 13.7 | -70.3 | -468.9 | 0.0 | -468.9 | -70.3 | 13.7 |
| 128.4 | -34.9 | -483.6 | -236.1 | -322.7 | -91.3 | -62.9 |
| 197.6 | 6.4 | -413.8 | -409.7 | -224.5 | -94.9 | -89.3 |
| 283.3 | 12.9 | -317.2 | -381.8 | -109.4 | -124.8 | -142.7 |

Figure 3-4. Imaginary Components

The center term of the real components is called the DC component. It is a measure of the average value of the image (7:15).

Because of the symmetry of the functions used to produce the FFT components, half of the components are duplicates. This symmetry can be seen by looking at the values in the arrays. As a result of this duplication of values, only half of the components are needed to produce a feature vector. The feature vector is formed by combining the DC term and 24 distinct real components with the 24 distinct imaginary components as shown in Figure 3-5 (7:16).

| | | | | | | |
|---|---|---|---|---|---|---|
| 248.3 | 197.6 | 128.4 | 13.7 | -62.9 | -89.3 | -142.7 |
| 12.9 | 6.4 | -34.9 | -70.3 | -91.3 | -94.9 | -124.8 |
| -317.2 | -413.8 | -483.6 | -468.9 | -322.7 | -224.5 | -109.4 |
| -381.8 | -409.7 | -236.1 | 666.4 | 149.6 | -97.2 | -142.1 |
| -113.5 | -65.6 | 298.5 | 607.0 | -28.7 | -116.6 | -193.2 |
| -76.2 | -32.2 | 380.5 | 448.4 | -188.6 | -117.2 | -13.6 |
| -48.4 | 19.6 | 173.1 | 221.7 | 110.2 | 31.6 | -42.5 |

Figure 3-5. Feature Vector Containing Real and Imaginary Components

# IV. Implementation

There were three stages of implementation, each consisting of modifications and additions to the AFRM developed by Lambert. In stage one, the AFRM was modified to use a new feature set. In stage two the original AFRM was modified to use a neural network for recognizing the feature vectors. After these two modifications were tested, stage three combined the two modifications to provide an AFRM which uses both concepts.

## 4.1 Stage 1: FaceDFT

The first stage of implementation was to modify the AFRM to use the 2DDFT to generate the feature set. Rather than developing a new 2DDFT algorithm, an already existing routine was used. A conventional 2DDFT subroutine was added to the AFRM. In addition to this new subroutine, modifications were made to several of the AFRM subroutines.

**4.1.1 Modifications to Gestalt.** The gestalt subroutine was modified to use the 2DDFT to calculate the feature set. It was decided that best results would be achieved if only brightness normalization is performed on the face before doing the 2DDFT. The 2DDFT is performed on each of the six windows. The lower two harmonics are filtered out as described in chapter 3, resulting in a 5 x 5 array of numbers for each of the six windows.

**4.1.2 Modification to Recognize.** The recognize subroutine was modified to use the new feature vector for making the recognition decision. The operations were modified to use 25 numbers from each window rather than 2 generated by the gestalt.

**4.1.3 Modification to the Data Base.** A new data base was developed to store the new feature vectors. The records in the data base consist of the name of the person, the picture number, and a 5 x 5 array of numbers from each of the six windows. each record is stored on 31 lines in the data base file. The name and face number are on the first line. Each of the six 5 x 5 arrays takes 5 lines. It was necessary to write new subroutines to read and write the database.

**4.1.4 Modification to Menu2.** Another modification was made to allow processing of the faces used by Lambert in his work. As much of Lambert's data as possible is being used in order to make a better comparison of the original AFRM with modified versions. To use this old data, it was necessary to display the individual faces back on the screen for reprocessing. An option was added to Menu2 for this purpose. When the option is taken, the user is prompted for the name of the .pic or .img file to be displayed.

## 4.2 Stage 2: FaceNet

The second stage consists of the modification of the original AFRM to use a neural network as the decision portion of the system. This began with the decision to use a back prop neural network. It was also decided that a separate program (back.c) would be developed to train the neural network, so that the training of the net can be done on a faster computer.

**4.2.1 Development of Back.c.** Back.c consists of calls to several subroutines. It is designed to work for any number of nodes in each of the three layers. The numbers are defined as constants at the beginning of the program. Following are the subroutines developed and their functions.

Readfile - This subroutine was copied from Face.c and is used to read in the training data base. It was modified to read the data into a 2D array rather than 2 1D arrays.

Init_Net - This subroutine takes care of initializing the network. All weights and thetas are set to small random values.

Set_Inputs - This subroutine receives a number as a parameter. This number is the record number of the training record to be used as input. The inputs of the net are then set to the values in the training record.

Calculate_Output - After the inputs have been set, this subroutine propagates the input values through the net to calculate the output values for each node.

Train_Net - This subroutine checks the outputs of the last layer of the net. It compares them to the desired values for the outputs. The errors are then used to modify the weights for the nodes. The error is then propagated back through the net to change the weights of the second and first layers.

Read_Net - This subroutine is used to read in the weights and thetas of a net, previously stored to a file.

Write_Net - This subroutine is used to save the weights and thetas of the net in a file for later use. The size of each layer and the number of inputs are also stored.

**4.2.2 Modification to Recognize.** Back.c was written and was executed to train and save a net. The AFRM was then modified to use this net. The modification consisted of adding the neural net subroutines Read_Net, Set_Inputs, and Calculate_Outputs to the AFRM. These subroutines were then used to modify the recognize subroutine of the AFRM. The neural net was tested using the test faces in the AFRM. Different network layer sizes were tested to find optimum performance.

### 4.3 Stage 3: FaceNetDFT

The final step in the software modification was to combine the two previous changes to the system. Each of the previous modification ideas was tested separately. After testing the two modifications were combined to create FaceNetDFT.c.

The combined modifications were similar to the individual changes. However, the neural net subroutines added to face.c had to be modified to use the DFT feature set instead of the original feature set. This also meant a change to Back.c, the program that trains the nets, resulting in the program BackDFT.c. Because of the length of time it takes to train a neural net, only the 25 numbers from the first window were used to train and test the DFT neural networks. An attempt was made to use 50 numbers as inputs, and it took over one week of CPU time to train the net.

# V. Results

This chapter presents the results of tests comparing Lambert's version of the AFRM, Face, with modified versions FaceNet, FaceDFT, and FaceNetDFT.

## 5.1 Face vs. FaceNet

A preliminary test was made comparing Face with several network configurations of FaceNet. The test was conducted using the database in the Face directory on SMV2A. It was later discovered that this was not the database used by Lambert, but the test remains valid. The database contained 14 test faces which were tested using Face, FaceNet with 100 layer1 nodes and 20 layer2 nodes, and FaceNet with 120 layer1 nodes and 12 layer2 nodes. The results are shown if Table 5-1. Both Face and FaceNet with 100x20 predicted 9 correct and 5 wrong, while FaceNet with 120x12 performed slightly worse, with 8 correct and 6 wrong.

Based on these preliminary results, work was started on FaceNetDFT, and a larger database was developed by building on Lambert's database. This final database, containing 24 people, was tested using Face and FaceNet. The results of this comparison are shown in Table 5-2.

## 5.2 Face vs. FaceDFT

Face was compared to FaceDFT using the 24 test faces in the final database. Best results were achieved from FaceDFT when the brightness normalized was used as opposed to the original picture or the contrast enhanced version. The results of the comparison are shown in table 5-3. They indicate slightly worse recognition by FaceDFT than by Face. Face had 16 correct and 8 wrong , while FaceDFT had 15 correct and 9

| Name | Face | FaceNet | FaceNet |
|---|---|---|---|
| MCalvo | Y | Y | Y |
| Remington | Y | Y | Y |
| JAdams | Y | Y | Y |
| LLambert | Y | Y | Y |
| MLambert | Y | SRogers | JAdams |
| DRuck | BHodges | CCrawford | CCrawford |
| SRogers | CCrawford | CCrawford | JHolt |
| DLambert | Y | Y | Y |
| BHodges | Y | Y | Y |
| JHolt | CCrawford | BHodges | BHodges |
| EC | DRuck | DRuck | DRuck |
| MKabrisky | Y | Y | Y |
| CCrawford | Y | Y | Y |
| Sander | DLambert | Y | BHodges |
| Total | Yes 9 | Yes 9 | Yes 8 |
| | No 5 | No 5 | No 6 |

Table 5-1. Preliminary Test of Face vs. FaceNet

wrong. Better recognition was expected from FaceDFT, so a search for a possible cause was made.

An examination was made of the faces that were not correctly recognized. Five of these faces were found to be somewhat tilted (See Figure 5-1). Two of these five were missed by Face and all five were missed by FaceDFT. This may be an indication that the 2DDFT is more sensitive to rotation than the "gestalt" calculation used in Face. If the results from the five bad faces (they do not fit the assumptions), were ignored then Face recognizes 13 correct and 6 wrong, and FaceDFT recognizes 15 correct and 4 wrong, which indicates a slightly better score for FaceDFT. Based on these results, work proceeded to develop FaceNetDFT.

| Name | Face | 120x20 FaceNet | 100x25 FaceNet |
|---|---|---|---|
| RMaple | n 2 | y | y |
| MKabrisky | n 2 | y | y |
| MLambert | y | gdawson | gdawson |
| LLambert | y | y | y |
| DLambert | n 2 | rmaple | y |
| SRogers | y | druck | mkabrisky |
| ECrawford | y | mkabrisky | mkabrisky |
| MMayo | y | gtarr | gtarr |
| JSillart | y | y | y |
| DBane | n 2 | y | y |
| DRuck | y | y | y |
| KCox | y | y | y |
| EFretheim | y | y | y |
| LRoberts | n 3 | dbane | ddoak |
| MDrylie | y | lroberts | y |
| GTarr | y | ppleva | llambert |
| CSabick | y | y | y |
| MOHair | y | y | y |
| PPleva | n 2 | csabick | csabick |
| DBridges | n 4 | y | y |
| DDoak | y | druck | y |
| GLorimor | n 6 | y | gdawson |
| RMorales | y | lroberts | y |
| GDawson | y | mdrylie | kcox |
| Total | yes 16<br>no 8 | yes 12<br>no 12 | yes 15<br>no 9 |

Table 5-2. Comparison of Face vs. FaceNet

## 5.3 Face vs. FaceNet DFT

The neural networks trained using the DFT input were limited to 25 inputs (one window) because of the computational limits of the computer used for training. An attempt was made to use 50 inputs for a network resulting in the use of over 1 week of cpu time and 3 weeks actual time for training.

| Name | Face | FaceDFT |
|---|---|---|
| RMaple | n 2 | y |
| MKabrisky* | n 2 | n 14 |
| MLambert | y | y |
| LLambert | y | y |
| DLambert | n 2 | n 2 |
| SRogers | y | y |
| ECrawford | y | n 4 |
| MMayo | y | y |
| JSillart | y | y |
| DBane | n 2 | y |
| DRuck* | y | n 3 |
| KCox | y | n 2 |
| EFretheim | y | y |
| LRoberts | n 3 | y |
| MDrylie* | y | n 7 |
| GTarr | y | y |
| CSabick | y | y |
| MOHair | y | n 2 |
| PPleva* | n 2 | n 5 |
| DBridges | n 4 | y |
| DDoak | y | y |
| GLorimor | n 6 | y |
| RMorales | y | y |
| GDawson* | y | n 14 |
| Totals | yes 16 | yes 15 |
| | no 8 | no 9 |
| Totals with no | yes 13 | yes 15 |
| bad pictures | no 6 | no 4 |

Table 5-3. Comparison of Face vs. FaceDFT

A network was therefore trained for each of the six windows. The results achieved by FaceNetDFT, using each of these networks, are shown in Table 5-4. Some windows performed better than others. The last column in the table indicates the results which would be achieved by using the name which was picked most by the six networks. This resulted

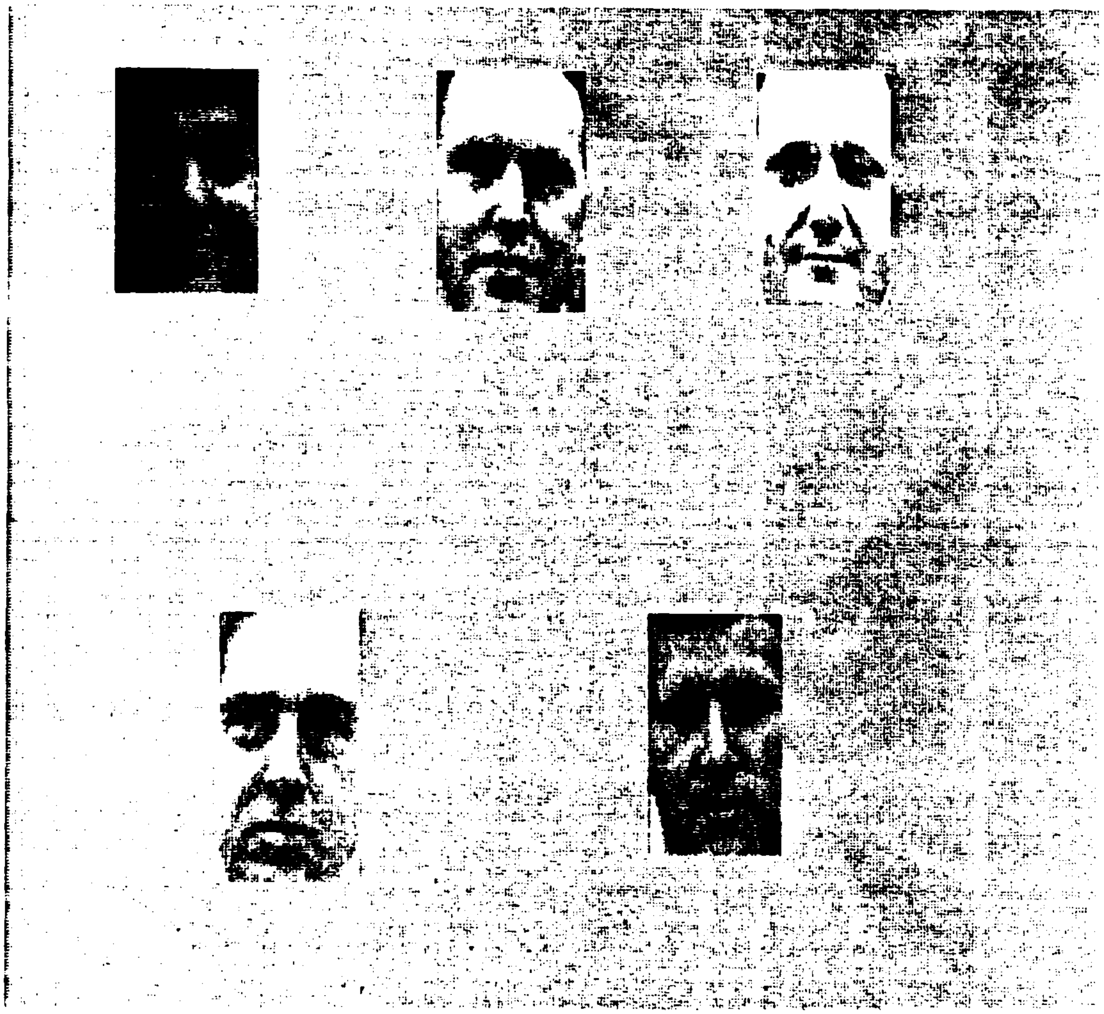

Figure 5-1. Bad Faces with Tilted Heads

in recognition of 18 correct, 4 wrong, and 2 ties. If the five questionable faces, discussed earlier, were not used, then the recognition score would be 18 correct, 0 wrong, and 1 tie. With or without these five faces, the total score for FaceNetDFT is better than that of Face.

| Name | Face | Win 1 | Win 2 | Win 3 | Win 4 | Win 5 | Win 6 | Total |
|---|---|---|---|---|---|---|---|---|
| RMaple | n 2 | y | y | y | y | y | y | y |
| MKabrisky* | n 2 | y | n | sroger | mmay | mlam | llamb | ? |
| MLambert | y | ccraw | y | y | y | y | y | y |
| LLambert | y | y | y | y | mmay | dbane | druck | y |
| DLambert | n 2 | y | mlam | y | y | y | y | y |
| SRogers | y | mlam | y | y | y | y | y | y |
| ECrawford | y | y | y | y | y | mlam | y | y |
| MMayo | y | y | ppleva | y | rmapl | mdryli | y | y |
| JSillart | y | y | y | y | rmapl | dlamb | y | y |
| DBane | n 2 | y | y | y | y | y | y | y |
| DRuck* | y | jsill | sroger | ddoak | dlamb | jsill | mmay | n |
| KCox | y | y | y | mdryli | y | y | y | y |
| EFretheim | y | y | y | y | y | y | y | y |
| LRoberts | n 3 | y | y | gtarr | rmoral | glorim | ecraw | y |
| MDrylie* | y | jsill | dlamb | jsill | rmapl | dlamb | jsill | n |
| GTarr | y | y | y | mohai | y | y | y | y |
| CSabick | y | rmoral | y | y | mlam | y | y | y |
| MOHair | y | y | y | n | ecraw | y | ecraw | y |
| PPleva* | n 2 | gdaw | dlamb | mdrylie | dlam | dlamb | mmay | n |
| DBridges | n 4 | y | y | y | efret | efret | y | y |
| DDoak | y | gdaw | ppleva | ppleva | y | gdaw | y | ? |
| GLorimor | n 6 | y | y | y | y | y | y | y |
| RMorales | y | y | glorim | dbane | glorim | y | y | y |
| GDawson* | y | dbane | rmapl | lrobert | jsill | druck | lrobert | n |
| Totals | yes 16 | yes 15 | yes 15 | yes 13 | yes 11 | yes 12 | yes 16 | yes 18 |
| | no 8 | no 9 | no 9 | no 11 | no 13 | no 12 | no 8 | no 4 |
| | | | | | | | | ? 2 |
| Totals with no bad pictures | yes 13 | yes 14 | yes 15 | yes 13 | yes 11 | yes 12 | yes 16 | yes 18 |
| | no 6 | no 5 | no 4 | no 6 | no 8 | no 7 | no 3 | no 0 |
| | | | | | | | | ? 1 |

Table 5-4. Comparison of Face vs. FaceNetDFT

## VI. Conclusions and Recommendations

### 6.1 Conclusions

The modification to use neural networks as the recognition portion of the AFRM proved to be worthwhile. The neural network provides recognition capability equivalent to that of the nearest neighbor system. In addition, once the network is trained, it provides the system with a constant recognition time (the time it takes the inputs to propagate through the network), independent of the number of faces in the database.

The use of the 2DDFT to generate features did not work as well as was hoped, however it did show some promise. The use of additional information from the 3rd harmonic of the DFT may provide better results. In addition, the use of an algorithm to eliminate problems caused by the tilted heads in some of the pictures may also improve the results. The use of the new feature information remains a valid avenue for future research.

### 6.2 Recommendations

This thesis effort concentrated only on a few of Lambert's recommendations for improving the AFRM. Lambert made many valid recommendations concerning the improvement of the AFRM's image processing and face location capabilities. Following are some of Lambert's recommendations, which remain valid:

1. Implement the processing of color images to increase the information available to the system. This may improve the separation of the face from the background, and possibly allow a better facial feature set.

2. Explore the use of binocular disparity techniques in the processing of images from a pair of cameras, to separate the face from the background.

3. Explore the limitations of the AFRM by training it with many more subjects. Develop methods to overcome these limitations.

Following are some additional recommendations which may be considered for future research:

1. Further explore the use of 2DDFTs in generating the feature set. Include the use of the 3rd Harmonic to provide more information. Explore methods to make the 2DDFT scale and rotation independent; for instance, preprocess the images in a Log z transform system as is known to be the case in the human visual system.

2. Verify the assumption that 4 images are sufficient to characterize a person in the database.

3. Explore the use of other neural network models in the recognition portion of the AFRM. Chapter 4 mentions several network models which may be used as classifiers.

4. Implement the training of the neural networks on a parallel processing (Encore) or a vector processing (Cray) computer. This should speed training time and allow the use of a larger number of inputs.

## Bibliography


1. Frisch, Bruce. "Connections: associative memory for computers," Aerospace America, 21: 12-13 (April 1987).

2. Fukushima, Kunihiko "Neocognitron: A Neural Network Model for a Mechanism of Visual Pattern Recognition" IEEE Transactions on SYSTEMS, Man, and Cybernetics, Vol SMC-13, No. 5: 826-834 (September/October 1983).

3. Gorman, Paul and Terrence J. Sejnowski "Analysis of Hidden Units in a Layered Network Trained to Classify Sonar Targets" Neural Networks, Vol 1: 75-89 (1988).

4. Heppenheimer, T. A. "Nerves of Silicon," Discover, 4: 70-79 (February 1988).

5. Lambert, Larry C. Evaluation and Enhancement of the AFIT Autonomous Face Recognition Machine, MS Thesis AFIT/GE/ENG/87D-35. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December 1987 (AD-A188819).

6. Lippmann, Richard P. "An Introduction to Computing with Neural Nets," IEEE ASSP Magazine, 73: 4-21 (April 1987).

7. O'Hair, Mark A. Whole Word Recognition Based On Low Freq Fourier Complex and Amplitude Spectra, MS Thesis AFIT/GEO/ENG/84D-4. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December 1984.

8. Rogers, Michael. "Mimicking the Human Mind," Newsweek, 102: 52-53 (July 20, 1987).

9. Routh, Richard L. Cortical Thought Theory: A Working Model of the Human Gestalt Mechanism, Ph.D. Dissertation AFIT/DS/EE/85-1. Air Force Institute of Technology (AU), Wright-Patterson AFB OH, July 1985.

10. Ruck, Dennis W. Multisensor Target Detection and Classification, MS Thesis AFIT/GE/ENG/87D-56. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December 1987.

11. Russel, Robert L. Jr. Performance of a working Face Recognition Machine Using Cortical Thought Theory, MS Thesis AFIT/GE/ENG/85D-37. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December 1985 (AD-A167781).

12. Smith, Edward J. Development of an Autonomous Face Recognition Machine, MS Thesis AFIT/GE/ENG/86D-36. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December 1986 (AD-A178852).

## Vita

Capt David D. Sander was born on [redacted] He grew up in western Pennsylvania where he attended Slippery Rock State College. He received a Bachelor of Science degree in Computer Science in May 1983. At the same time he received his commission through the ROTC program.

He served as a computer systems analyst at the Foreign Technology Division at Wright-Patterson AFB until entering the School of Engineering, Air Force Institute of Technology, in June 1987.

[redacted]

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

## REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE

3. DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release; distribution unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S): AFIT/GCS/ENG/88D-19

5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION: School of Engineering

6b. OFFICE SYMBOL (If applicable): AFIT/ENG

7a. NAME OF MONITORING ORGANIZATION

6c. ADDRESS (City, State, and ZIP Code): Air Force Institute of Technology, Wright-Patterson AFB OH 45433-6583

7b. ADDRESS (City, State, and ZIP Code)

8a. NAME OF FUNDING / SPONSORING ORGANIZATION

8b. OFFICE SYMBOL (If applicable)

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

8c. ADDRESS (City, State, and ZIP Code)

10. SOURCE OF FUNDING NUMBERS

| PROGRAM ELEMENT NO. | PROJECT NO. | TASK NO | WORK UNIT ACCESSION NO. |
|---|---|---|---|
| | | | |

11. TITLE (Include Security Classification): ENHANCED AUTONOMOUS FACE RECOGNITION MACHINE, VOL I

12. PERSONAL AUTHOR(S): David D. Sander, B.S., Capt, USAF

13a. TYPE OF REPORT: MS thesis

13b. TIME COVERED: FROM ______ TO ______

14. DATE OF REPORT (Year, Month, Day): 1988 December

15. PAGE COUNT

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

| FIELD | GROUP | SUB-GROUP |
|---|---|---|
| 12 | 09 01 | |
| 06 | 04 | |

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Pattern Recognition; Face; Discrete Fourier Transforms

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Thesis Advisor: Matthew Kabrisky, PhD
Professor of Electrical Engineering

Approved for release in Accordance with AFR 190-1
[signature]
12 Jan 1989

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT: ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL: Matthew Kabrisky, PhD

22b. TELEPHONE (Include Area Code): (513) 255-5276

22c. OFFICE SYMBOL: AFIT/ENG

DD Form 1473, JUN 86 — Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

## Abstract

This thesis continues work on the Autonomous Face Recognition Machine developed at AFIT in 1985. There were two major changes made to the system. The set of features extracted from the face for use in the recognition process, was changed. A higher dimensioned vector taken from the two-dimensional Discrete Fourier Transform of the face, was used in hope of increasing the separation of templates stored in the data base. Further research is needed to determine whether this change is beneficial to the system. The second change was to the decision rule used in recognition. The decision making portion of the system was replaced by a back propagation neural network. While providing equivalent recognition capability, this change provides a constant recognition time independent of the number of subjects trained into the system.